CE 211
Mathematics for Engineers

Lecture 4

Discrete Random Variables

Previously on Mathematics for Engineers

To summarize, a probability space consists of three components

- A sample space Ω which is the set of all outcomes
- A set of events $\mathcal{F}$
- A probability measure or a function $\mathbb{P} : \mathcal{F} \to [0, 1]$

The probability measure must satisfy the following three axioms.

Axioms

1. For every event $A \in \mathcal{F}$, $\mathbb{P}(A) \geq 0$
2. $\mathbb{P}(\Omega) = 1$
3. If $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint events, i.e., $A_i \cap A_j = \emptyset \, \forall i, j$, then

$$\mathbb{P}\big(\cup_{i=1}^{\infty} A_i \big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

Previously on Mathematics for Engineers

Definition (Conditional Probability)

If A and B are two events and if $\mathbb{P}(B) > 0$, $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$

Definition (Law of Total Probability)

Suppose $A_1, \ldots, A_n$ represents a partition of the sample space Ω and $\mathbb{P}(A_i) > 0 \, \forall \, i = 1, \ldots, n$. Then, for any event B

$$\mathbb{P}(B) = \mathbb{P}(A_1)\mathbb{P}(B|A_1) + \ldots + \mathbb{P}(A_n)\mathbb{P}(B|A_n)$$

Theorem (Bayes' Theorem)

Suppose $A_1, \ldots, A_n$ represents a partition of the sample space Ω and $\mathbb{P}(A_i) > 0 \, \forall \, i = 1, \ldots, n$. Then, for any event B with $\mathbb{P}(B) > 0$

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_i)\mathbb{P}(B|A_i)}{\mathbb{P}(B)}$$

Previously on Mathematics for Engineers

Definition (Independence)

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, two events A and B in $\mathcal{F}$ are said to be independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Claim

If two events A and B are independent, $\mathbb{P}(A|B) = \mathbb{P}(A)$ *and* $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Definition (Independence of n Events)

The $A_1, A_2, \ldots, A_n$ are said to be independent if

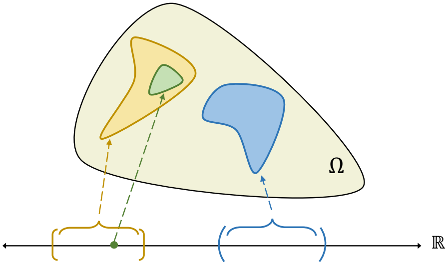
$$\mathbb{P}\left(\cap_{I \subset \{1,\ldots,n\}} A_i\right) = \Pi_{i \in I} \mathbb{P}(A_i)$$

Previously on Mathematics for Engineers

A random variable is an alternate way of constructing events. Defining random variables allows us to translate events of interest into probabilities more easily.

Definition (Random Variable)

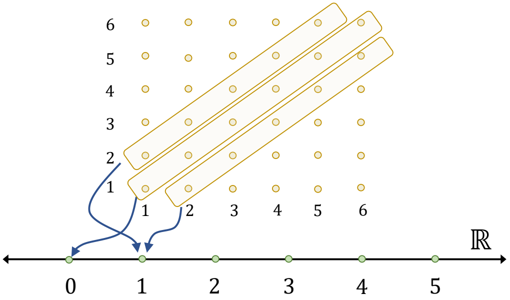
A real-valued random variable is a function or mapping $X : \Omega \to \mathbb{R}$ such that for all $S \subset \mathbb{R}$, $X^{-1}(S) \in \mathcal{F}$.



*Technically, there are some restrictions on S just like valid events, but we'll ignore those details.

Previously on Mathematics for Engineers

Instead, if we define X as the absolute value of difference in the numbers on the dice. What is the event corresponding to $X = 7$? $X = 1$?



Previously on Mathematics for Engineers

Note the probability measure is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$ where you can think of $\mathcal{F}$ as 2^{Ω}, whereas the random variable X is another function $X : \Omega \to \mathbb{R}$.

As seen in the previous examples, for subsets $S \subset \mathbb{R}$, we can find an event $A \in \mathcal{F}$ such that $X^{-1}(S) = A = \{\omega \in \Omega | X(\omega) \in S\}$.

Hence, the following probabilities are the same

$$\mathbb{P}(X \in S) = \mathbb{P}(X^{-1}(S)) = \mathbb{P}(A) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in S\})$$

Be careful to not write $X(A)$ and $\mathbb{P}(S)$, where $A \in \mathcal{F}$ and $S \subset \mathbb{R}$ (unless of course $\Omega = \mathbb{R}$).

Previously on Mathematics for Engineers

Definition (Probability Mass Function)

The probability mass function (PMF) of a random variable X represents the probability of each outcome. It is denoted as $p_X(x)$ and is defined as

$$p_X(x) = \mathbb{P}(X = x)$$

Definition (Cumulative Distribution Function)

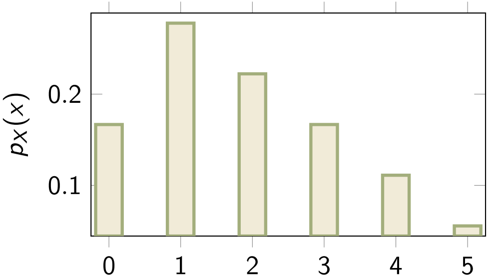
The cumulative distribution function (CDF) of a random variable X is denoted by $F_X(x)$ and is defined as

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{x' \leq x} p_X(x')$$

Previously on Mathematics for Engineers

In the problem of throwing two dice, what is the PMF of the random variable defined as the absolute value of the difference of the numbers on the faces.

x	$p_X(x)$
0	$6/36$
1	$10/36$
2	$8/36$
3	$6/36$
4	$4/36$
5	$2/36$

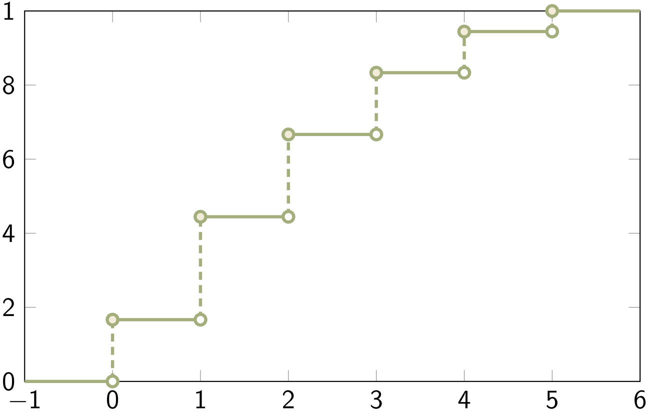


Previously on Mathematics for Engineers

The CDF function is always non-decreasing and right continuous. It also satisfies

$$\lim_{x \to -\infty} F_X(x) = 0 \qquad \lim_{x \to +\infty} F_X(x) = 1$$

Further, for discrete random variables the CDF is a step function.



Lecture Outline

Lecture Outline

The Problem

The Problem

St. Petersburg Paradox

Suppose we play a game in which I toss an unbiased coin until it lands on Tails. So the outcomes of this experiment could be T, HT, HHT, $HHHT$, and so on.

We perform the experiment once and suppose I promise to give you ₹ 2^{n+1} where n is the number of heads that we will see, but first you need to pay me to participate in this game.

How much are you willing to pay to enter this game? How many of you wouldn't mind paying ₹100? ₹1,000?

Expectation and Variance

Expectation and Variance

Introduction

Imagine we perform an experiment repeatedly. Each outcome can be mapped to real number using a random variable and we can compute the average of all realizations. This notion of average is also called the expectation and is formally defined as

Definition (Expectation)

The expected value of a random variable X is denoted by $\mathbb{E}(X)$ or μ_X and is defined as

$$\mathbb{E}(X) = \sum_{x \in R_X} x p_X(x)$$

- Imagine a random variable X in the coin toss experiment which takes $+1$ for H and -1 for T. What is its expectation?
- What is the expectation of the random variable representing the absolute value of the difference for the two-dice experiment?

We can extend this definition of expectation to functions of random variables $f(X)$ (which is just a composite function of f and X)

Definition (Expectation of Functions)

The expectation of a function of random variable $f(X)$ is denoted by $\mathbb{E}(f(X))$ or $\mu_{f(X)}$ and is defined as

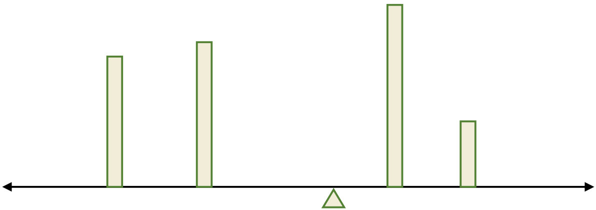
$$\mathbb{E}(f(X)) = \sum_{x \in R_X} f(x) p_X(x)$$

In the coin toss experiment, if you are rewarded according to function $f(X) = X^2$, what is the expected reward from the experiment?

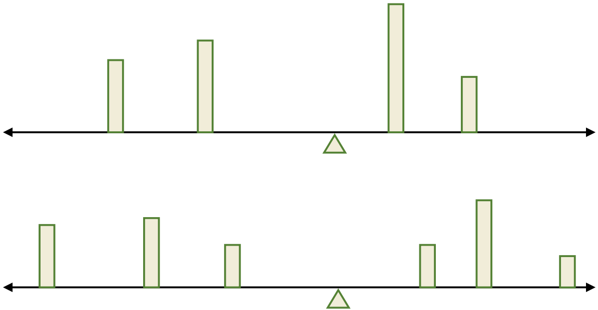
Claim (Linearity of Expectation)

If a and b are constants, $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$

You could imagine expectation to be a point which balances the probability masses (i.e., the torque at that point is 0). Why?



The following PMFs have the same expectation but what is different about them?



Expectation and Variance

Variance

The extend of dispersion or spread around the mean is captured by variance.

Definition (Variance)

The variance of a random variable X is denoted by $V(X)$, $\mathrm{Var}(X)$, or σ_X^2 and is defined as

$$V(X) = \mathbb{E}\left((X - \mu_X)^2\right) = \sum_{x \in R_X} (x - \mu_X)^2 p_X(x)$$

The term $\sigma_X = \sqrt{V(X)}$ is also called the standard deviation of X.

Prove the following results using the above definition:

Claim

- $V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$
- *If a and b are constants $V(aX + b) = a^2 V(X)$*

Binomial and Bernoulli Distribution

Binomial and Bernoulli Distribution

Motivating Example

Consider tossing a unbiased coin repeatedly 3 times. What is the probability that we see exactly 2 heads.

The possible outcomes are

$$HHH, HHT, HTT, HTH, THH, THT, TTH, TTT$$

If we define the random variable X as the number of heads. What is

- The support of X?
- The probability mass function of X?
- $\mathbb{E}(X)$ and $V(X)$?

x	$p_X(x)$
0	1/8
1	3/8
2	3/8
3	1/8

Binomial and Bernoulli Distribution

Motivating Example

What if we instead use a biased coin and the probability of heads is p and tails is $1-p$?

Again, the possible outcomes are

$$HHH, HHT, HTT, HTH, THH, THT, TTH, TTT$$

but they are not equally likely. Define the random variable X as the number of heads and use the independence of tosses and determine

- The support of X?
- The probability mass function of X?
- $\mathbb{E}(X)$ and $V(X)$?

x	$p_X(x)$
0	$(1-p)^3$
1	$3p(1-p)^2$
2	$3p^2(1-p)$
3	p^3

Binomial and Bernoulli Distribution

Assumptions

Let's now generalize these results under the following assumptions:

- The experiment consists of n independent repeated trials
- Each trial produces two outcomes defined as a success and a failure
- The probability of success in each trial is p

Definition (Binomial Random Variable)

In this context, a Binomial random variable X is defined as the number of successes in n trials and is denoted by $X \sim \text{Bin}(n, p)$.

The values of n and p are also called parameters of the distribution.

Binomial and Bernoulli Distribution

Probability Mass Function

One can choose x successes from n trials in $\binom{n}{x}$ ways and the probability with which this happens is p^x. The remaining trials must produce failures which happens with probability $(1-p)^{n-x}$.

Claim (PMF of Binomial Random Variable)

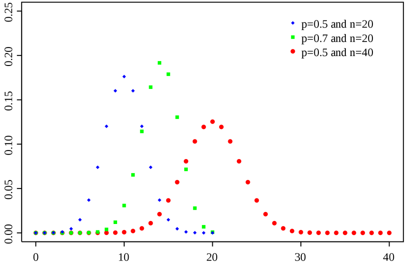
Suppose $X \sim Bin(n, p)$, then

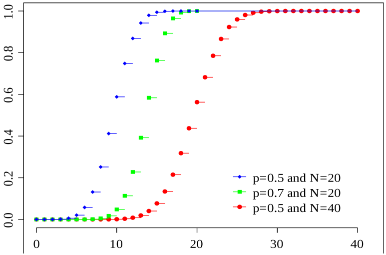
$$\mathbb{P}(X = x) = p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \textit{if } x = 0, 1, \ldots, n \\ 0 & \textit{otherwise} \end{cases}$$

- The support of X is $0, 1, \ldots, n$.
- The above PMF is valid since its sum across all realizations of X is 1. (Why?)

Binomial and Bernoulli Distribution

PMF and CDF





Claim (Expectation and Variance)

If $X \sim Bin(n,p)$, $\mathbb{E}(X) = np$ and $V(X) = np(1-p)$.

Proof.

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} n \frac{(n-1)!}{(x-1)!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{x=0}^{n-1} np \frac{(n-1)!}{(x)!(n-x-1)!} p^x (1-p)^{n-x-1} = np
\end{aligned}$$

$V(X)$ (Exercise) ∎

Binomial and Bernoulli Distribution

Special Cases

The Binomial Distribution for the special case when $n = 1$ is called **Bernoulli Distribution**.

For this reason, each trial in the Binomial experiment is also called a Bernoulli trial.

Suppose an unbiased coin is tossed 8 times. What is the probability of getting at most 3 heads.

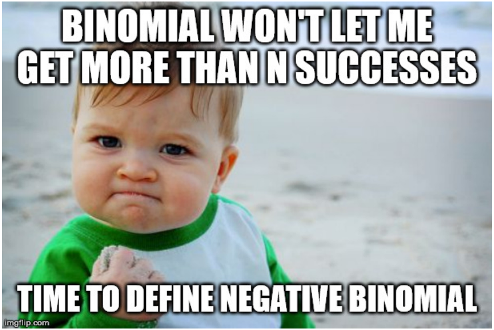
Lecture Outline

Negative Binomial and Geometric Distribution

Negative Binomial and Geometric Distribution

Introduction

Suppose instead we perform Bernoulli trials until r successes are observed. What are the outcomes of the experiment?



Negative Binomial and Geometric Distribution

Introduction

Definition (Negative Binomial Random Variable)

Let X be the number of failures before r successes. It is said to be Negative Binomial distributed and is denoted as $X \sim NB(r, p)$

The parameters of this distribution are r and p.

Another way of defining a negative binomial random variable involves counting the number of trials until we observe r successes. What is the support of this random variable?

Since we stop after r successes, the last trial must be a success. Hence, we must get $r-1$ successes in $x+r-1$ trials. Thus, using Binomial distribution,

Claim (PMF of Negative Binomial Random Variable)

Suppose $X \sim NB(r,p)$

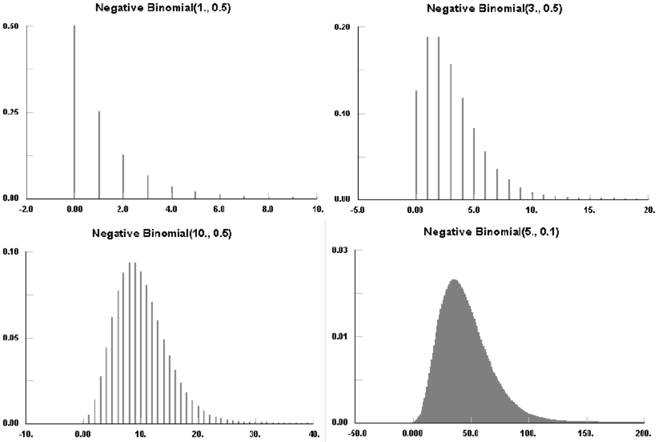
$$\mathbb{P}(X=x) = p_X(x) = \begin{cases} \binom{x+r-1}{r-1} p^r (1-p)^x & \text{if } x = 0,1,2,\ldots \\ 0 & \text{otherwise} \end{cases}$$

- The support of X is $0,1,2,\ldots$.
- The above PMF is valid since its sum across all realizations of X is 1. To prove this, use the binomial theorem for negative exponents. Suppose $|x| < y$,

$$(x+y)^{-n} = \sum_{k=0}^{\infty} (-1)^k \binom{n+k-1}{k} x^k y^{-n-k}$$

Negative Binomial and Geometric Distribution

PMF and CDF



Negative Binomial and Geometric Distribution

Expectation and Variance

Claim

If $X \sim NB(r,p)$, *then* $\mathbb{E}(X) = \frac{r(1-p)}{p}$ *and* $V(X) = \frac{r(1-p)}{p^2}$

Proof.

$$
\begin{aligned}
\mathbb{E}(X) &= \sum_{x=0}^{\infty} x \binom{x+r-1}{r-1} p^r (1-p)^x \\
&= \sum_{x=1}^{\infty} x \frac{(x+r-1)!}{x!(r-1)!} p^r (1-p)^x \\
&= r(1-p)p^r \sum_{x=1}^{\infty} \frac{(x+r-1)!}{(x-1)!r!} (1-p)^{x-1} \\
&= r(1-p)p^r \sum_{x=0}^{\infty} \frac{(x+(r+1)-1)!}{x!((r+1)-1)!} (p-1)^x(-1)^x = r(1-p)p^r p^{-(r+1)}
\end{aligned}
$$

$V(X)$ (Exercise) ∎

Negative Binomial and Geometric Distribution

Special Cases

A geometric distribution is a special case of negative binomial in which we are interested in the number of failures until just 1 success, i.e., $NB(1, p)$

The PMF for the geometric distribution follows a geometric sequence.

Suppose an unbiased coin is tossed till we see 3 heads. What is the probability that we see at least 3 tails.

Poisson Distribution

Poisson Distribution

Introduction and Assumptions

The Possion random variable is used to count the number of random events in a time period.

For example,

- The number of accidents that occur on a highway in an year.
- The number of customers served by a teller at a bank.

It is assumed that occurrences of events are independent of each other. Also, we assume that two or more events cannot happen simultaneously. Further, the average rate of occurrence of events is known and assumed constant.

Poisson Distribution

Probability Mass Function

Poisson Distribution can be viewed as a limit of the binomial distribution. For instance, suppose the average number of occurrences is known to be λ. Let the time period of interest be divided into n intervals and assume that at most one occurrence can happen in each interval.

The success probability of an occurrence in each period is therefore λ/n. From the binomial distribution, the probability of observing x occurrences is

$$\binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

Taking limits as $n \to \infty$

$$\lim_{n\to\infty} \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} = \frac{\lambda^x}{x!} \lim_{n\to\infty} \frac{n!}{(n-x)!}\left(\frac{1}{n^x}\right)\left(1-\frac{\lambda}{n}\right)^{n-x}$$

Poisson Distribution

Probability Mass Function

$$= \frac{\lambda^x}{x!} \lim_{n\to\infty} \frac{n(n-1)\dots(n-x+1)}{n^x} \left(1 - \frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\lambda^x}{x!} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n}$$

$$= \frac{e^{-\lambda}\lambda^x}{x!}$$

Definition

PMF of a Poisson distributed random variable with parameter $\lambda > 0$ is

$$\mathbb{P}(X = x) = p_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Poisson Distribution

Expectation and Variance

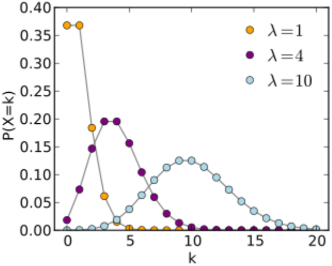
Claim

Suppose $X \sim Pois(\lambda)$, $\mathbb{E}(X) = V(X) = \lambda$

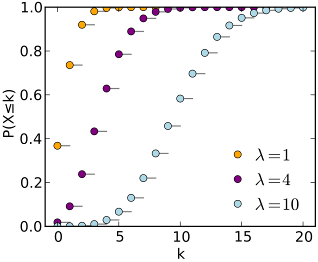
Proof.

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^{x}}{x!} \\
&= \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{(x-1)!} \\
&= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\
&= \lambda e^{-\lambda} e^{\lambda}
\end{aligned}$$

$V(X)$ (Exercise) ■



P(X=k)
k
λ=1
λ=4
λ=10



P(X≤k)
k
λ=1
λ=4
λ=10

Poisson Distribution

Quickercise

Assume that on an average 4 students check out books at the library every hour. What is the probability that at least 4 students check out books in a 4 hour time window.

A Solution

A Solution

St. Petersburg Paradox

- Construct the PMF of the reward.
- Bernoulli and Cramer in the 1700s proposed the concept of diminishing marginal utility of money. If you have more money, every extra rupee is less valuable. Utility functions are concave. (Imagine log functions. What is your new expected value?)
- There is also risk associated with the experiment. What is the variance of the total reward?

Your Moment of Zen